

**This Page is Inserted by IFW Indexing and Scanning  
Operations and is not part of the Official Record**

**BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- BLACK BORDERS**
- IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- FADED TEXT OR DRAWING**
- BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- SKEWED/SLANTED IMAGES**
- COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- GRAY SCALE DOCUMENTS**
- LINES OR MARKS ON ORIGINAL DOCUMENT**
- REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- OTHER:** \_\_\_\_\_

**IMAGES ARE BEST AVAILABLE COPY.**

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.

PRCA-77071-V  
09/492,249

日本特許庁  
PATENT OFFICE  
JAPANESE GOVERNMENT

別紙添付の書類に記載されている事項は下記の出願書類に記載されて  
いる事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed  
with this Office.

出願年月日  
Date of Application:

1998年12月25日

出願番号  
Application Number:

平成10年特許願第369589号

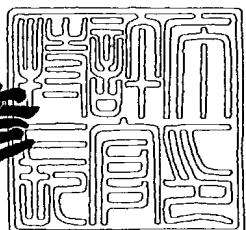
出願人  
Applicant(s):

株式会社リコー

1999年10月22日

特許庁長官  
Commissioner,  
Patent Office

近藤 隆彦



出証番号 出証特平11-3072806

【書類名】 特許願

【整理番号】 9806051

【提出日】 平成10年12月25日

【あて先】 特許庁長官殿

【国際特許分類】 G06F 17/27

【発明の名称】 文書処理装置、文書処理方法およびその方法をコンピュータに実行させるプログラムを記録したコンピュータ読み取り可能な記録媒体

【請求項の数】 13

【発明者】

【住所又は居所】 東京都大田区中馬込1丁目3番6号 株式会社リコー内

【氏名】 宮地 達生

【発明者】

【住所又は居所】 東京都大田区中馬込1丁目3番6号 株式会社リコー内

【氏名】 嶋田 敦夫

【発明者】

【住所又は居所】 東京都大田区中馬込1丁目3番6号 株式会社リコー内

【氏名】 剣持 栄治

【発明者】

【住所又は居所】 東京都大田区中馬込1丁目3番6号 株式会社リコー内

【氏名】 山崎 真湖人

【発明者】

【住所又は居所】 東京都大田区中馬込1丁目3番6号 株式会社リコー内

【氏名】 武谷 一寿

【発明者】

【住所又は居所】 東京都大田区中馬込1丁目3番6号 株式会社リコー内

【氏名】 長束 哲郎

【特許出願人】

【識別番号】 000006747

【氏名又は名称】 株式会社リコー

【代理人】

【識別番号】 100104190

【弁理士】

【氏名又は名称】 酒井 昭徳

【手数料の表示】

【予納台帳番号】 041759

【納付金額】 21,000円

【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【包括委任状番号】 9810808

【ブルーフの要否】 要

【書類名】 明細書

【発明の名称】 文書処理装置、文書処理方法およびその方法をコンピュータに実行させるプログラムを記録したコンピュータ読み取り可能な記録媒体

【特許請求の範囲】

【請求項1】 入力された複数の文書データを所定の形式で表示または印刷するために出力する文書処理装置において、

入力された文書データを記憶する文書記憶手段と、

前記文書記憶手段により記憶された文書データの全部または一部を選択する選択手段と、

前記選択手段により選択された文書データの全部または一部から文字列の特徴に関するデータを抽出する特徴抽出手段と、

前記特徴抽出手段により抽出された文字列の特徴に関するデータに基づいて前記文書データの全部または一部を加工処理する加工処理手段と、

前記加工処理手段により加工処理された文書データの全部または一部を出力する出力手段と、

を備えたことを特徴とする文書処理装置。

【請求項2】 前記出力手段は、

前記加工処理手段により加工処理された文書データの全部または一部の内容に基づいて複数の項目値を設定する項目値設定手段と、

前記項目値設定手段により設定された項目値ごとに前記文書データの全部または一部を集計する集計手段と、

を備え、

前記文書データの全部または一部を、項目値を少なくとも1つの軸とする表形式に展開して出力することを特徴とする請求項1に記載の文書処理装置。

【請求項3】 前記出力手段は、さらに、前記加工処理手段により加工処理された文書データの全部または一部を、前記加工処理手段により加工処理される前の文書データの全部または一部とともに出力することを特徴とする請求項1または2に記載の文書処理装置。

【請求項4】 前記文書記憶手段は、さらに、前記加工処理手段により加工

処理された文書データの全部または一部を記憶することを特徴とする請求項1～3のいずれか一つに記載の文書処理装置。

【請求項5】 前記選択手段は、さらに、前記出力手段により出力された文書データの全部または一部を選択することを特徴とする請求項1～4のいずれか一つに記載の文書処理装置。

【請求項6】 前記文書記憶手段は、さらに、前記加工処理の内容に関するデータを記憶することを特徴とする請求項1～5のいずれか一つに記載の文書処理装置。

【請求項7】 入力された複数の文書データを所定の形式で表示または印刷するために出力する文書処理方法において、

入力された文書データを記憶する文書記憶工程と、

前記文書記憶工程により記憶された文書データの全部または一部を選択する選択工程と、

前記選択工程により選択された文書データの全部または一部から文字列の特徴に関するデータを抽出する特徴抽出工程と、

前記特徴抽出工程により抽出された文字列の特徴に関するデータに基づいて前記文書データの全部または一部を加工処理する加工処理工程と、

前記加工処理工程により加工処理された文書データの全部または一部を出力する出力工程と、

を含んだことを特徴とする文書処理方法。

【請求項8】 前記出力工程は、

前記加工処理工程により加工処理された文書データの全部または一部の内容に基づいて複数の項目値を設定する項目値設定工程と、

前記項目値設定工程により設定された項目値ごとに前記文書データの全部または一部を集計する集計工程と、

を含み、

前記文書データの全部または一部を、項目値を少なくとも1つの軸とする表形式に展開して出力することを特徴とする請求項7に記載の文書処理方法。

【請求項9】 前記出力工程は、さらに、前記加工処理工程により加工処理

された文書データの全部または一部を、前記加工処理工程により加工処理される前の文書データの全部または一部とともに出力することを特徴とする請求項7または8に記載の文書処理方法。

**【請求項10】** 前記文書記憶工程は、さらに、前記加工処理工程により加工処理された文書データの全部または一部を記憶することを特徴とする請求項7～9のいずれか一つに記載の文書処理方法。

**【請求項11】** 前記選択工程は、さらに、前記出力工程により出力された文書データの全部または一部を選択することを特徴とする請求項7～10のいずれか一つに記載の文書処理方法。

**【請求項12】** 前記文書記憶工程は、さらに、前記加工処理の内容に関するデータを記憶することを特徴とする請求項7～11のいずれか一つに記載の文書処理方法。

**【請求項13】** 前記請求項7～12のいずれか一つに記載された方法をコンピュータに実行させるプログラムを記録したことを特徴とするコンピュータ読み取り可能な記録媒体。

#### 【発明の詳細な説明】

##### 【0001】

##### 【発明の属する技術分野】

この発明は、入力された複数の文書データを所定の形式で表示または印刷するために出力する文書処理装置、文書処理方法およびその方法をコンピュータに実行させるプログラムを記録したコンピュータ読み取り可能な記録媒体に関する。

##### 【0002】

##### 【従来の技術】

近年、さまざまな文書分類装置や文書検索装置が開発されている。文書分類装置としては、たとえば、特開平7-114572号公報に記載されているように、文書から自動的に単語の特徴ベクトルを抽出し、その特徴ベクトルをもとに文書分類することで、意味的な異なりを用いた自動分類を可能にするものがある。

##### 【0003】

##### 【発明が解決しようとする課題】

しかしながら、上記従来技術の文書分類装置や文書検索装置等の文書処理装置は、単に文書を分類する、あるいは文書を検索する機能を有するのみで、その結果を用いてさらなる分析をおこない、文書群に内在する隠れた情報の解析をおこなうことについては何ら考慮がされておらず、文書群に内在する隠れた情報の解析は別の解析装置をもちいておこなわなければならないという問題点があった。

#### 【0004】

情報分析をおこなう操作者が分類作業や検索作業をおこなうのは、これらの作業において、結果は目的なのではなく、単に情報分析作業の途中経過にすぎないからである。通常は、その後、さらに結果を把握しやすくするために、元の文書に含まれる情報を最大限に活用し、結果の並べ替えをおこなったり、集計・統計処理を施したり、結果をもとに表の形式にまとめたり、さらにはグラフ化したりというようなさまざまな処理を繰り返しおこない、意味ある情報分析結果を導き出す必要がある。

#### 【0005】

また、数値データを対象とする情報の分析作業において、表計算ソフトウェアが用いられる場合があるが、表計算ソフトウェアは、元来、数値データの取り扱いを意図して開発されたものであり、文字データ、特に文書の意味に係わるような分析作業においては十分な効果を発揮することはできなかった。

#### 【0006】

この発明は、上述した従来例による問題点を解消するため、文書の意味に係わるような分析作業において、単に分類作業や検索作業などを固定された機能としておこない、その結果を出力するのではなく、情報分析作業全般にわたる支援をおこなうことができる文書処理装置、文書処理方法およびその方法をコンピュータに実行させるプログラムを記録したコンピュータ読み取り可能な記録媒体を提供することを目的とする。

#### 【0007】

##### 【課題を解決するための手段】

上述した課題を解決し、目的を達成するため、請求項1の発明に係る文書処理装置は、入力された複数の文書データを所定の形式で表示または印刷するために

出力する文書処理装置において、入力された文書データを記憶する文書記憶手段と、前記文書記憶手段により記憶された文書データの全部または一部を選択する選択手段と、前記選択手段により選択された文書データの全部または一部から文字列の特徴に関するデータを抽出する特徴抽出手段と、前記特徴抽出手段により抽出された文字列の特徴に関するデータに基づいて前記文書データの全部または一部を加工処理する加工処理手段と、前記加工処理手段により加工処理された文書データの全部または一部を出力する出力手段と、を備えたことを特徴とする。

#### 【0008】

また、請求項2の発明に係る文書処理装置は、請求項1の発明において、前記出力手段が、前記加工処理手段により加工処理された文書データの全部または一部の内容に基づいて複数の項目値を設定する項目値設定手段と、前記項目値設定手段により設定された項目値ごとに前記文書データの全部または一部を集計する集計手段と、を備え、前記文書データの全部または一部を、項目値を少なくとも1つの軸とする表形式に展開して出力することを特徴とする。

#### 【0009】

また、請求項3の発明に係る文書処理装置は、請求項1または2の発明において、前記出力手段が、さらに、前記加工処理手段により加工処理された文書データの全部または一部を、前記加工処理手段により加工処理される前の文書データの全部または一部とともに出力することを特徴とする。

#### 【0010】

また、請求項4の発明に係る文書処理装置は、請求項1～3の発明において、前記文書記憶手段が、さらに、前記加工処理手段により加工処理された文書データの全部または一部を記憶することを特徴とする。

#### 【0011】

また、請求項5の発明に係る文書処理装置は、請求項1～4の発明において、前記選択手段が、さらに、前記出力手段により出力された文書データの全部または一部を選択することを特徴とする。

#### 【0012】

また、請求項6の発明に係る文書処理装置は、請求項1～5の発明において、

前記文書記憶手段が、さらに、前記加工処理の内容に関するデータを記憶することを特徴とする。

#### 【0013】

また、請求項7の発明に係る文書処理方法は、入力された複数の文書データを所定の形式で表示または印刷するために出力する文書処理方法において、入力された文書データを記憶する文書記憶工程と、前記文書記憶工程により記憶された文書データの全部または一部を選択する選択工程と、前記選択工程により選択された文書データの全部または一部から文字列の特徴に関するデータを抽出する特徴抽出工程と、前記特徴抽出工程により抽出された文字列の特徴に関するデータに基づいて前記文書データの全部または一部を加工処理する加工処理工程と、前記加工処理工程により加工処理された文書データの全部または一部を出力する出力工程と、を含んだことを特徴とする。

#### 【0014】

また、請求項8の発明に係る文書処理方法は、請求項7の発明において、前記出力工程が、前記加工処理工程により加工処理された文書データの全部または一部の内容に基づいて複数の項目値を設定する項目値設定工程と、前記項目値設定工程により設定された項目値ごとに前記文書データの全部または一部を集計する集計工程と、を含み、前記文書データの全部または一部を、項目値を少なくとも1つの軸とする表形式に展開して出力することを特徴とする。

#### 【0015】

また、請求項9の発明に係る文書処理方法は、請求項7または8の発明において、前記出力工程が、さらに、前記加工処理工程により加工処理された文書データの全部または一部を、前記加工処理工程により加工処理される前の文書データの全部または一部とともに出力することを特徴とする。

#### 【0016】

また、請求項10発明に係る文書処理方法は、請求項7~9の発明において、前記文書記憶工程が、さらに、前記加工処理工程により加工処理された文書データの全部または一部を記憶することを特徴とする。

#### 【0017】

また、請求項11の発明に係る文書処理方法は、請求項7～10の発明において、前記選択工程が、さらに、前記出力工程により出力された文書データの全部または一部を選択することを特徴とする。

#### 【0018】

また、請求項12の発明に係る文書処理方法は、請求項7～11の発明において、前記文書記憶工程が、さらに、前記加工処理の内容に関するデータを記憶することを特徴とする。

#### 【0019】

また、請求項13の発明に係る記憶媒体は、請求項7～12に記載された方法をコンピュータに実行させるプログラムを記録したことで、そのプログラムを機械読み取り可能となり、これによって、請求項7～12の動作をコンピュータによって実現することが可能である。

#### 【0020】

##### 【発明の実施の形態】

以下に添付図面を参照して、この発明に係る文書処理装置、文書処理方法およびその方法をコンピュータに実行させるプログラムを記録したコンピュータ読み取り可能な記録媒体の好適な実施の形態を詳細に説明する。

#### 【0021】

まず、この発明の実施の形態による文書処理装置を構成する情報処理システム全体のハードウェア構成を説明する。図1は、本実施の形態による文書処理装置を構成する情報処理システム全体のハードウェア構成を示す説明図である。

#### 【0022】

図1において、本実施の形態による文書処理装置を構成する情報処理システムは、サーバー／クライアント方式で構成されている。すなわち、サーバー101と複数のクライアント102がネットワーク103によって接続されている。クライアント102は、分類データ等の加工データの生成、サーバー101への指示、分類結果等の加工処理結果の表示などを起こす。一方、クライアント102からの指示にしたがって、サーバー101は文書（テキスト）分類等の加工処理を膨大な数値演算により起こす、その処理の結果をクライアント102へ送

る。

### 【0023】

分類処理の場合、より具体的には、サーバー101においては、テキスト分類処理（前処理、クラスタリング処理）がおこなわれ、クライアント102においては、分類データ生成、処理実行指示、テキスト分類結果表示等がおこなわれる。サーバー101における処理は、上述のように、「前処理」と「分類処理」の二つに分かれており、その処理はデータによっては非常に負荷が大きくなる。したがって、サーバー101は「前処理」と「分類処理」がそれぞれ一つずつしか処理をおこなわないようにマネージャプロセスが処理受付リストを作成して管理する。

### 【0024】

また、サーバー101とクライアント102との間のデータのやりとりはファイル共有という方法を用いる。すなわち、分類処理等の加工処理に用いるファイルをサーバー101上の共有フォルダに作成することにより両者はデータのやりとりをおこなう。したがって、クライアント102からはサーバー101の共有フォルダをネットワーク共有して利用することが可能である。

### 【0025】

つぎに、サーバー101およびクライアント102のハードウェア構成について説明する。図2は、本実施の形態による文書処理装置を構成する情報処理システムにおけるサーバー101をハードウェア的に示す説明図である。サーバー101は、たとえばワークステーション(WS)等が用いられる。

### 【0026】

図2において、201はサーバー101全体を制御するCPUを、202はブートプログラム等を記憶したROMを、203はCPU201のワークエリアとして使用されるRAM203を、204は通信回線205を介してネットワーク103に接続され、そのネットワーク103と内部のインターフェイスを司るインターフェイス(I/F)を、206はデータを記憶するディスク装置を示している。200は上記各部を結合させるためのバスを示している。

### 【0027】

そのほか、文書情報、画像情報、機能情報等を表示するディスプレイ208や、データを入力するためのキーボード209およびマウス210等が同様に接続されていてもよい。さらに、ディスク装置206には、クライアント102との間のデータのやりとりをするための共有フォルダ207が設けられている。

#### 【0028】

また、図3は、本実施の形態による文書処理装置を構成する情報処理システムにおけるクライアント102をハードウェア的に示す説明図である。クライアント102は、たとえばパーソナルコンピュータ（PC）等が用いられる。

#### 【0029】

図3において、301はシステム全体を制御するCPUを、302はブートプログラム等を記憶したROMを、303はCPU301のワークエリアとして使用されるRAMを、304はCPU301の制御にしたがってHD（ハードディスク）305に対するデータのリード／ライトを制御するHDD（ハードディスクドライブ）を、305はHDD304の制御で書き込まれたデータを記憶するHDを、306はCPU301の制御にしたがってFD（フロッピーディスク）307に対するデータのリード／ライトを制御するFDD（フロッピーディスクドライブ）を、307はFDD306の制御で書き込まれたデータを記憶する着脱自在のFDを、308はドキュメント、画像、機能情報等を表示するディスプレイをそれぞれ示している。

#### 【0030】

また、309は通信回線310を介してネットワーク103に接続され、そのネットワーク103と内部のインターフェイスを司るインターフェイス（I/F）を、311は文字、数値、各種指示等の入力のためのキーを備えたキーボードを、312はカーソルの移動や範囲選択、あるいは表示画面に表示されたアイコンやボタンの押下やウインドウの移動やサイズの変更等をおこなうマウスを、313はOCR（Optical Character Reader）機能を備えた画像を光学的に読み取るスキヤナを、314は分類結果を含むデータの内容等を印刷するプリンタを、315は上記各部を結合するためのバスをそれぞれ示している。また、HD305にはワープロソフト等のアプリケーションソフト3

16が記憶されている。

#### 【0031】

つぎに、本実施の形態による文書処理装置の機能的構成について説明する。図4は、本実施の形態による文書処理装置の構成を機能的に示すブロック図である。図4において、文書処理装置は、入力部401と、文書記憶部402と、選択部403と、特徴抽出部404と、加工処理部405と、出力部406を含む構成である。

#### 【0032】

入力部401、文書記憶部402、選択部403、特徴抽出部404、加工処理部405、出力部406は、ROM202または302、RAM203または303、あるいはディスク装置306またはハードディスク316等の記録媒体に記録されたプログラムに記載された命令にしたがってCPU201または301等が命令処理を実行することにより、各部の機能を実現する。

#### 【0033】

入力部401は、文書データを入力するものであり、たとえば、キーボード209または311、スキャナ313、OCR機能を備えたスキャナ313、またはネットワーク103を経由して文書や文書群を得ることができるI/F204または309等である。また、入力部401は、上記以外に、文書データを取得することができるものであれば、それらのすべてを含む。たとえば、文書データがデータベース化されている場合に、そのデータベースが記録された媒体を本実施の形態の文書処理装置に組み入れた場合も文書データの入力とする。

#### 【0034】

ここで、文書とは、自然言語で記述された一つ以上の文の集まりであり、文字、文字列、数値等から構成されており、それらの意味のあるまとまりを1つの文書とする。また、複数の文書の集まりを、文書群とする。

#### 【0035】

文書は1つあるいは複数の項目から構成されている。項目は、項目名と、項目値から構成されている。項目名は項目の内容を示すラベルであり、文書に含まれていても含まれていなくてもよい。項目値は項目の実際の内容である。図5は、

本実施の形態による文書処理装置の項目名と項目値の関係を示す説明図である。たとえば、1つの特許公報は1つの文書であり、特許公報を項目名と項目値によって表現すると、図5のようになる。

#### 【0036】

入力部401によって取得された文書あるいは文書群は、それぞれの文書に一意な文書IDが付与され、文書記憶部402により記憶される。図6は、本実施の形態による文書処理装置の文書記憶部402に記憶された文書のデータ構造を示す説明図である。各項目名あるいは項目値は、文書記憶部402のセル、すなわち1つの記憶単位に収納される。

#### 【0037】

図6においては、1つのセルは3つの記憶領域から構成されており、第1番目の記憶領域601には次のセルの文書記憶部402上の位置（番地）が記憶されている。第2番目の記憶領域602には、セルの属性値が記憶されている。

#### 【0038】

セルの属性値としては、たとえば、「0」が「空」、「1」が「数値」、「2」が文字列・・・というように設定することができる。第3番目の記憶領域603には、セルの実際の内容、すなわち、項目名あるいは項目値等が格納される領域の先頭位置が記憶されている。

#### 【0039】

セルの順序の並び替えや、セルの追加・削除は、第1番目の記憶領域601に記憶された次のセルの位置を変更することにより、容易に実現することができる。また、セルの実際の内容は、セルの構造とは異なる別の領域に記憶されているので、たとえば、項目を変更した結果、あらかじめ確保された領域では取り切れなくなった場合には、セルの構造自体には影響なく、別途大きな領域を確保してそこに項目値を記憶し、第3番目に記憶された記憶領域603の先頭位置を変更するだけでよい。

#### 【0040】

図7は、本実施の形態による文書処理装置の文書記憶部402に記憶された文書の別のデータ構造を示す説明図である。図7において、1つのセルは2つの記

憶領域を使用している。第1番目の記憶領域701には、セル属性値が記憶されている。第2番目の記憶領域702には、セルの実際の内容、すなわち項目名あるいは項目値などが格納される領域の先頭位置が記憶されている。

#### 【0041】

つぎのセルは、文書記憶部402上でとなり合うつぎの記憶領域に記憶されている。このデータ構造では、セルの順序の並び替え、セルの追加・削除が発生した場合には、記憶内容の移動操作が必要となる。

#### 【0042】

文書記憶部402は、通常高速に情報を扱える半導体メモリで構成されるが、磁気ディスクあるいは光ディスク等で構成される補助記憶装置を含んでいてよい。

#### 【0043】

文書記憶部402に記憶された文書あるいは文書群は、出力部406により表示される。本実施の形態においては、出力部406は、CRTディスプレイ、液晶ディスプレイ等から構成される。出力部406は、文書記憶部402に記憶された文書あるいは文書群の内容をセルと順次たどって読み出し、表の形式で表示または印刷する。

#### 【0044】

また、出力部406は、表の形式で表示または印刷されたデータに基づいてグラフを描画するグラフ描画部407を含んでいてよい。グラフ描画部407は、文書記憶部402に記憶された文書あるいは文書群の項目値に対して利用者が設定した領域の内容を読み出し、利用者の指示により棒グラフ、円グラフ、折れ線グラフ等のグラフを描画し、表示または印刷する。

#### 【0045】

出力部406は、入力部401による操作に関する表示、たとえば、操作メニューやマウスポインタ、カーソルの表示等もおこなう。また、処理結果を印刷するためのプリンタ等の印刷装置を含んでいてよい。

#### 【0046】

選択部403は、入力部401による操作者の指示により、出力部406の表

示上で選択された領域のデータを文書記憶部402から読み出し、特徴抽出部404へ送る。選択部403の選択方法について、図8～図10を用いて説明する。

#### 【0047】

図8～図10は、本実施の形態による文書処理装置の出力部406による画面表示の例、具体的には、自動車の故障状況の内容が表示された画面表示の例を示す説明図である。図8において、画面表示には、文書ID番号を示す「番号」欄801、故障情報を受け付けた日付を示す「受付日」欄802、故障情報を受け付けた営業所を示す「営業所」欄803、故障情報の対象となった自動車の車種を示す「車種」欄804、故障情報対象となった自動車の年式を示す「年式」欄805、故障状況の内容を示す「内容」欄806が表示される。

#### 【0048】

図9において、選択領域901は、矩形で囲まれ、表示色が変更されている部分であり、図10においても同様に、選択領域1001は、矩形で囲まれ、表示色が変更されている部分である。

#### 【0049】

選択部403が選択する領域としては、図9に示すように、画面上の列の一部であってもよいし、また、図10に示すように項目名を選択した場合はその項目名に属する項目値全部が選択されるようにしてもよい。なお、本実施の形態では、文字列の属性をもつ領域のみ選択可能とする。

#### 【0050】

つぎに、特徴抽出部404によりおこなわれる抽出処理の内容について説明する。選択部403により選択された項目値は、特徴抽出部404によりその項目値の特徴が抽出される。図11は、本実施の形態による文書処理装置の特徴抽出部404によりおこなわれる抽出処理の内容の一覧を示す説明図である。

#### 【0051】

図11において、抽出処理には、対象とする文字列に含まれる単語、その単語の単語数、単語の文字数、単語のそれぞれの出現回数...等がある。これらの抽出処理は、規則音声合成装置や自動翻訳装置等の一般的に用いられている形態

素解析技術あるいは構文解析技術等の自然言語処理技術を用いて実現する。

#### 【0052】

つぎに、加工処理部405によりおこなわれる加工処理の内容について説明する。特徴抽出部404により抽出処理された特徴量に対して、加工処理部405により加工処理が施される。図12は、本実施の形態による文書処理装置の加工処理部405によりおこなわれる加工処理の内容の一覧を示す説明図である。

#### 【0053】

加工処理には、同一の特徴量ごと分類する「分類処理」、所定の特徴量を検索する「検索処理」、特徴量の内容ごとに並べ替えをおこなう「並べ替え処理」、特徴量の代表値を抽出する「代表値抽出処理」、特徴量のうちの最大値を抽出する「最大値抽出処理」、特徴量のうち最小値を抽出する「最小値抽出処理」、特徴量を算術する「算術処理」等がある。

#### 【0054】

特徴抽出部404によりおこなわる特徴量の抽出処理の内容と、加工処理部405によりおこなわれる抽出された特徴量の加工処理の内容の組み合わせは、おのおの操作者が選択できるようにすることができる。また、効果の高い組み合わせをあらかじめ設定して、その設定された組み合わせを操作者に提供するようにしてもよい。

#### 【0055】

加工処理部405により加工処理された処理結果は、加工処理部405内の加工処理結果保持部408に保持される。加工処理結果保持部408に保持された加工処理結果は、出力部406により出力される。出力部406は、加工処理結果保持部408から内容を読み出し、画像表示や印刷出力をおこなう。

#### 【0056】

ここで、特徴抽出部404により抽出される特徴（量）として、項目値に含まれる単語それぞれの出現回数を選択し、加工処理部405によりおこなわれる加工処理として、分類処理を選択した場合について説明する。

#### 【0057】

一般的に、2つの文書があり、それら2つの文書を構成する単語の出現頻度が

等しい場合、それら2つの文書の意味は似通っていると考えることができる。すなわち、ある文書での単語の出現回数は、その文書の意味に関係の深い特徴量であると考えることができる。したがって、単語の出現回数を特徴量として、複数の文書を分類した場合、それぞれの分類カテゴリーには意味の近い文書が所属する考えることができる。

#### 【0058】

選択部403により選択された1つあるいは複数の項目値は、特徴抽出部404に含まれる解析部409によって項目値ごとに形態素解析等の自然言語解析をおこない、単語に分割される。また、それぞれの単語には、その単語の品詞情報も付与される。出現した単語のうち、名詞であるものに対して一意な単語IDを付与し、1つの項目値および選択部403により選択されたすべての項目値に対する単語IDごとの出現回数を計数する。

#### 【0059】

特徴抽出部404に含まれる特徴ベクトル生成部410は、計数された出現回数に基づいて個々の項目値の特徴（量）を示す項目値特徴ベクトルを生成する。たとえば、選択部403により選択された項目値が、

「騒音が大きい」

「塗装が変色する」

「オーバーヒートが起こる」

「塗装がはげる」

「バッテリーが上がる」

「排気が黒い」

であった場合、各項目の特徴ベクトルは、図13に示すようになる。また、図14には、単語とその単語IDごとの出現回数を示す。

#### 【0060】

すなわち、

「騒音が大きい」 : {1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0}

, 0, 0}

「塗装が変色する」 : {0, 1, 0, 1, 1, 1, 0, 0, 0, 0, 0, 0}

, 0, 0 }

「オーバーヒートが起こる」 : {0, 1, 0, 0, 0, 0, 1, 1, 0, 0, 0  
, 0, 0}

「塗装がはげる」 : {0, 1, 0, 1, 0, 0, 0, 0, 1, 0, 0  
, 0, 0}

「バッテリーが上がる」 : {0, 1, 0, 0, 0, 0, 0, 0, 0, 1, 1  
, 0, 0}

「排気が黒い」 : {0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0  
, 1, 1}

という特徴ベクトルが得られる。

#### 【0061】

この項目値の特徴ベクトルは、特徴抽出部404からの出力として加工処理部405へ送られる。加工処理部405においては、項目値の特徴ベクトルをもついて、分類処理をおこなう。分類処理は、まず、おののおののベクトル間の距離を計算することにより求める。距離の尺度には、たとえば内積をもちいることができる。

#### 【0062】

距離を計算した後、距離の近いものをまとめる処理をおこなう。この処理には、たとえばベクトルの集合をその距離に応じてK個のベクトルの集合に分類するK-means法をもちいることができる。それぞれのベクトルの分類が完了したら、ベクトルに対応する項目値に対してどの分類に属するかの番号、すなわち、クラスタ番号と、項目値に対応する文書IDを付与し、加工処理部405の出力とし、出力部406により出力表示をおこなう。

#### 【0063】

図15は、クラスタ番号1501を表示した画面の表示例を示す。クラスタ番号が同一番号である文書、たとえば、番号「1」および「6」の文書のクラスタ番号が「5」であり、両者が同一の分類に属することを示している。

#### 【0064】

つぎに、請求項2の発明においてクロス表を出力する場合について説明する。

入力部401により、分析対象とする文書群を読み込んだ後、操作者は分類処理をおこなう対象となる項目名、クロス表の横軸あるいは縦軸となる項目名、いくつに分類をおこなうかの分類数を指示する。

#### 【0065】

図16はクロス表作成のための指示画面である。図16において、指示画面1600は、処理対象項目名入力欄1601と、軸となる項目名入力欄1602と、縦軸指定ボタン1603と、横軸指定ボタン1604と、分類数入力欄1605とから構成される。

#### 【0066】

処理対象項目名入力欄1601には、処理対象となる項目名を入力する。キーボード209等から入力するあるいは処理対象となる項目候補を表示させその中からマウス210等により選択することにより項目名を入力することができる。また、軸となる項目名入力欄1602には、軸となる項目名を入力する。入力の方法は、処理対象項目名入力欄1601への入力の方法と同様である。

#### 【0067】

縦軸指定ボタン1603および横軸指定ボタン1604は、軸となる項目を縦軸に表示させるか横軸に表示させるかを指定するためのボタンである。また、分類数入力欄1605には、いくつに分類するかその分類数を入力する。入力の方法としては、キーボード209等から数字を入力するあるいは分類数候補を表示させその中からマウス210等により選択することにより分類数を入力するようにもよい。

#### 【0068】

図16においては、処理対象項目名入力欄1601には「内容」が、軸となる項目名入力欄1602には「車種」が、また、横軸指定ボタン1604がチェックされ、分類数入力欄1605には「50」が入力され、これにより、文書群の中の「内容」に基づいて、「50(個)」に分類され、クロス表の横軸に「車種」を表示するという指示がなされていることがわかる。

#### 【0069】

クロス表作成の指示がおこなわれることにより、分類処理が実行され、その結

果がクロス表で表示される。図17および図18は、分類処理の結果が表示されたクロス表を示す図である。図17において、クロス表1700は、縦軸に分類を示す「クラスタ1」、「クラスタ2」...が表示され、横軸に車種を示す「ABC1600」、「ABC1800」...が表示される。

#### 【0070】

表の縦軸、すなわち各行は、分類処理により生成されたクラスタに対応する。各行の第1欄には、分類処理終了時には既定値としてクラスタ番号を示す文字列が入っている。表の横軸、すなわち各欄には、文書群の項目「車種」に含まれる文字列が重複することなく表示される。行「クラスタ1」の各セルには、クラスタ1に分類された文書のうち、項目「車種」の値がその欄の車種と一致するものの数が表示される。

#### 【0071】

ここで、数を表示する代わりに、セルの色の濃淡や、セルを塗りつぶす面積により数の大きさを表現するようにしてもよい。また、表の最右欄および最下欄には、該当する行、欄の合計が表示される。

#### 【0072】

図18において、クロス表1700のあるセルにマウスポインタ1800を移動させ、マウス210のマウスボタンを押下する、あるいはキーボード209のカーソルキー操作によりカーソルを移動させ、特定キーを押下すると、そのセルの近傍に内容表示画面1801が表示されることにより、該当する文書の項目「内容」が表示される。

#### 【0073】

内容表示画面1801には、セル内のデータ数、表示項目、セル情報、および、各データにおける表示項目の内容が表示される。マウスポインタ1800により指定されたセルにおいては、データ数：「4」、表示項目：「内容」、セル情報：「ABC2000-クラスタ1」、表示項目の内容として「内容」の4つの内容である「排気が黒い、排気が黒い、...」が表示される。これにより、マウスポインタを所望のセルに移動させてマウスボタンを押下するという簡易な操作により、セルの内容を認識することができる。

## 【0074】

また、内容表示画面1801に表示される項目は、設定操作により変更することが可能であり、すべての項目を表示させることもでき、また、項目を選択して表示させることもできる。

## 【0075】

各行の第1欄には、分類処理終了時には既定値としてクラスタ番号を示す文字列が入っているが、操作者により、この欄の書き換えをすることができる。たとえば、上記の操作によってセルの内容を確認した後、「クラスタ1」を「排気の問題」と書き換えることができる。これにより、情報内容の把握がより容易になる。

## 【0076】

また、分類終了時に既定値としてクラスタ番号を示す文字列を入れるのではなく、そのクラスタの特徴を示す文字列を抽出し、セルに入れることも可能である。たとえば、クラスタ1に含まれる文書の項目「内容」から、もっとも頻度が高く出現する文や単語を抽出する事により実現する。

## 【0077】

図18においては、クラスタ1には「排気が黒い」あるいは「排気」等の単語が入れられる。このように、操作者は簡易な操作により文書全体の分布状態を把握するだけでなく、必要に応じて個々の文書の内容をも詳細に知ることができる。

## 【0078】

つぎに、クロス表を作成するための出力部406の詳細な構成の内容について説明する。図19は、本実施の形態による文書処理装置の出力部406の詳細な構成を示すブロック図である。出力部406は、グラフ描画部407のほかに、項目値選定部1901、集計部1902とから構成され、集計部はさらに実際に表示する内容に対応した記憶領域を持つ表保持部1903を備えている。

## 【0079】

項目値選定部1901は、操作者がクロス表の1つの軸として指定した項目名（軸項目名）に対して、文書記憶部402に記憶された文書データから、項目値

を順次読み出し、重複のない項目値の集合を作成する。また、集計部1902は、表保持部1903の項目値に対応する領域に数値を加算することにより文書の集計をおこなう。

#### 【0080】

つぎに、クロス表の出力手順について説明する。図20は、本実施の形態による文書処理装置のクロス表の出力手順を示すフローチャートである。図20のフローチャートにおいて、まず、集計に先立ち、表保持部1903の内容を初期化する（ステップS2001）。

#### 【0081】

つぎに、項目値設定部1901により作成された項目値を、表の項目値ラベルに相当する部分に当てはめ（ステップS2002）、クラスタ番号を表す文字列を、クラスタ番号に相当する部分に当てはめる（ステップS2003）。

#### 【0082】

つぎに、加工処理結果保持部408に保持された項目値に対応する文書IDについて、文書記憶部402に記憶された対応する文書を参照し、その軸項目名に對応する項目値を決定する（ステップS2004）。その後、表保持部1903の対応する領域の内容に1を加算する（ステップS2005）。

#### 【0083】

すべての項目値について上記処理をおこなったか否かを判断し（ステップS2006）、すべての項目値について上記処理がおこなわれていない場合（ステップS2006否定）は、ステップS2004へ移行し、ステップS2004～S2006の処理を繰り返しおこなう。

#### 【0084】

ステップS2006において、すべての項目値について上記処理がおこなわれた場合（ステップS2006肯定）は、最右列に表示するための行の合計を計算し（ステップS2007）、あわせて、最下行に表示するための欄の合計を計算する（ステップS2008）。

#### 【0085】

その後、表保持部1903に構成された表を、順次読み出して出力し（ステッ

PS2009)、すべての処理を終了する。

#### 【0086】

なお、加工処理部405から出力されたデータを、文書記憶部402に送り、文書記憶部402に他のデータとともに記憶するように構成してもよい。文書記憶部402に記憶された加工処理部405から出力されたデータは、出力部406によって表の新たな列として表示することができる。また、表の既存の列を消去し、消去した列へ上書きするようにしてもよい。

#### 【0087】

この構成では、処理の結果である加工処理部405から出力されたデータは、文書記憶部402において、今回の加工処理の対象とならなかった他のデータと対等に取り扱うことができ、その後の分析作業等で、もともとの入力データに存在していたか、分析作業の途中で加工処理によって生成されたのかを区別することなく、加工処理の対象として選択することが可能である。

#### 【0088】

したがって、データの性質や、おこないたい情報分析作業の内容に応じて柔軟に加工処理対象と加工処理内容を選択する事ができるので、多彩で高度な情報分析作業が可能となる。

#### 【0089】

また、加工処理部405への入力データとして、特徴抽出部404から出力されたデータだけではなく、選択部403により選択されたデータも含めることができる。これにより、文字列の特徴抽出を必要としないデータや、加工処理結果の数値に対しても更なる加工処理を施すことができるので、より多彩で高度な情報分析が可能となる。

#### 【0090】

図21～図24は、本実施の形態による文書処理装置の出力部406による画面表示の別の例を示す説明図である。図21において、「番号」、「受付日」、「営業所」、「車種」、「年式」、「内容」の他に、分類処理により得られた結果である「クラスタ番号」2101が表示されている。

#### 【0091】

さらに、図21においては、選択部403により「クラスタ番号」2101が選択されており、「クラスタ番号」2101に関するデータが反転表示されている。選択された「クラスタ番号」2101をキーとして、加工処理部405により並べ替え処理をおこなうよう指示をする。

#### 【0092】

並べ替え処理の指示により、並べ替え処理がおこなわれた結果を表示しているのが図22である。図22においては、「クラスタ番号」が「1」のものが集まって表示されるように並べ替えられ、それに続き、「クラスタ番号」が「2」のものが集まって表示されるように並び替えられる。

#### 【0093】

具体的には、「クラスタ番号」が「1」である「番号」が「2」、「11」、「15」、「23」、「35」、「54」、「63」、「73」、「82」の順で並べ替えられ、それに続き「クラスタ番号」が「2」である「番号」が「14」、「18」、「22」、「27」、「37」、...が表示されていることがわかる。

#### 【0094】

つぎに、項目「車種」の欄で、「クラスタ番号」が「1」に属するものを選択する。図23においては、項目「車種」の欄で、「クラスタ番号」が「1」に属するものが選択され、その選択領域2301が反転表示されていることを示している。このように、すでに「クラスタ番号」により並べ替えがおこなわれており、同一クラスタに属するものが集まって表示されているので、画面上の連続した領域として容易に選択することができる。

#### 【0095】

つぎに、選択領域2301について車種別の発生頻度の棒グラフを表示させたのが、図24である。図24において、棒グラフ表示領域2401には、選択領域2301によって選択された「クラスタ番号」が「1」である9つの文書が選択され、その9つの文書を車種別に棒グラフ化したものが表示される。

#### 【0096】

このように、加工処理の対象を柔軟かつ容易に選択でき、選択された対象につ

いて多様な加工処理をおこなうことができ、また、その加工処理結果も次回の加工処理の対象とすることができますので、高度な情報分析作業が可能となる。

#### 【0097】

このように、分類等の文字列の特徴量を抽出して、その特徴量をもちいておこなう加工処理を実施した後に多種の加工処理をおこなう例を示したが、事前に多種の処理をおこなうことができるようにしてよい。

#### 【0098】

たとえば、「車種」の項目を選択し、これをキーとして並べ替えをおこなった後、集まつたある車種、たとえば、「ABC1600」に対して分類処理をおこなうこともできる。また、入力部401により入力された文書が誤字等の誤りを含んでいる場合、分類等の文字列の特徴量を抽出して、その特徴量をもちいて加工処理をおこなう前に、たとえば、文字列の検索・置換処理をおこなって、誤字を一括して修正し、より好適な結果が得られるようにデータを整えることもできる。

#### 【0099】

図25は、本実施の形態による文書処理装置の文書記憶部402の詳細な構成を示すブロック図である。図25において、文書記憶部402は、設定値記憶部2501および設定値送受信部2502を含んでいる。設定値記憶部2501には、文書を分類する際の分類数等の分類情報記憶部2503をはじめとするさまざまな設定値、すなわち文書処理装置の動作に必要な設定値に関する情報を記憶する記憶部を備えている。これにより設定値に関する情報は、文書情報とともに記憶することができる。

#### 【0100】

また、設定値送受信部2502は、設定値記憶部2501によって記憶された設定値に関する情報を他の情報処理装置へ送信する。また、設定値送受信部2502は、他の情報処理装置からの設定値に関する情報を受信する。設定値送受信部2502により受信された設定値に関する情報は、設定値記憶部2501によって記憶される。

#### 【0101】

記憶された設定値に関する情報は、後に文書を再度読み込んだときに同時に読み込まれ設定値記憶部2501に記憶される。この設定値に関する情報は操作者が所定の操作をすることにより参照することができたり、以後の処理の際に、再利用することができる。これにより、設定値に関する情報を文書とともに保存・管理することができる。設定値に関する情報の紛失を防ぎ、好適な設定値を後に再利用することができる。

#### 【0102】

図26～図28は、本実施の形態による文書処理装置の出力部406による画面表示の別の例を示す説明図である。図26において、まず、操作者が分類をおこなうべき対象である「内容」を表示画面上で選択する。それにより選択領域2601が反転表示される。つぎに、メニュー・バー2603から、分類処理ボタン2603を選択すると、分類処理に必要な分類数、すなわち、対象をいくつに分類するかについての問い合わせ画面2604が表示される。

#### 【0103】

操作者が問い合わせ画面2604において分類数を入力すると、この分類数に関する情報が文書記憶部402に記憶される。図26においては、分類数として「50」が入力されたことを示している。

#### 【0104】

その後、操作者が情報分析作業を完了して、メニュー・バー2603のファイルボタン2605の選択によりポップアップする図示を省略する保存ボタンを押下すると、文書記憶部402により、操作者が指示したファイル名が付与され、文書の情報、分類結果とともに記憶される。

#### 【0105】

図27において、分類結果を表示する欄2701にマウスポインタ2702を移動させ、マウスボタンを押下すると、その分類をおこなうことにもちいた分類に関する情報および分類設定値に関する情報を表示する分類情報表示画面2703が表示される。これにより、用いた設定値の関連付けが容易に把握することができる。

#### 【0106】

分類情報表示画面2703には、たとえば、分類に関する情報として分類がおこなわれた日時に関する情報を示す「分類日時」、分類の対象となった文書数に関する情報を示す「分類対象数」等が表示され、また、分類設定値に関する情報として、いくつに分類したかを示す「分類数」、どの品詞に基づいて分類をしたかを示す「分類品詞」等が表示される。

#### 【0107】

分類処理を実行するたびに新規な表が作成される。図28は、分類結果1を得た後、再度分類処理がおこなわれ、分類結果2が表示された状態を示している。分類結果1を再度表示させたい場合は、画面左下部のラベル上の選択領域2801へマウスポインタを移動させ、マウスボタンを押下する。これにより、分類結果1が再度表示される。その後、分類結果2を再度表示させる場合も同様の操作によりおこなうことができる。

#### 【0108】

また、図28において、各分類処理の実行にもちいた設定値に関する情報が対応する表の所定の表示領域2802に表示される。この表示領域2802は、分類結果の表示を隠さないように表示させることができ、また、その表示位置を移動することもできる。これにより、分類結果と、それにもちいた設定値の関連付けが容易に把握できる。

#### 【0109】

つぎに、本実施の形態における文書処理装置の文書処理の一連の手順について説明する。図29は、本実施の形態による文書処理装置の文書処理の一連の手順を示すフローチャートである。

#### 【0110】

図29のフローチャートにおいて、まず、文書データが文書処理装置に入力されたか否かを判断する（ステップS2901）。ここで、文書データが入力されるのを待って、文書データが入力された場合（ステップS2901肯定）は、入力された文書データを記憶する（ステップS2902）。なお、ステップS2901およびS2902の各ステップは、文書の入力があるごとに他のステップとは独自におこなわれるようにしてよい。

## 【0111】

つぎに、記憶された文書データの全部または一部が選択されたか否かを判断する（ステップS2903）。ここで、文書データの全部または一部が選択されるのを待って、選択された場合（ステップS2903肯定）は、選択された文書データの全部または一部の文字列の特徴に関するデータの抽出をおこなう（ステップS2904）。

## 【0112】

その後、ステップS2904において、抽出された文字列の特徴に関するデータに基づいて、分類処理等、所定の加工処理をおこなう（ステップS2905）。つづいて、ステップS2905において加工処理がおこなわれたデータを、表形式に展開する等の出力処理をおこなう（ステップS2906）。

## 【0113】

さらに、ステップS2905において加工処理されてデータを元の文書データに関連付けして記憶する（ステップS2907）。また、加工処理の設定値等の加工処理の内容に関するデータもあわせて記憶する（ステップS2908）。

## 【0114】

その後、ステップS2905において加工処理されたデータの全部または一部が選択されたか否かを判断し（ステップS2908）、選択された場合（ステップS2908肯定）は、ステップS2904へ移行し、以後、ステップS2904～S2909の処理を繰り返しおこなう。一方、ステップS2909において、加工処理されたデータの全部または一部が選択されなかった場合（ステップS2909否定）は、すべての処理を終了する。

## 【0115】

## 【発明の効果】

以上説明したように、請求項1の発明によれば、入力された文書データを記憶する文書記憶手段と、前記文書記憶手段により記憶された文書データの全部または一部を選択する選択手段と、前記選択手段により選択された文書データの全部または一部から文字列の特徴に関するデータを抽出する特徴抽出手段と、前記特徴抽出手段により抽出された文字列の特徴に関するデータに基づいて前記文書デ

ータの全部または一部を加工処理する加工処理手段と、前記加工処理手段により加工処理された文書データの全部または一部を出力する出力手段とを備えるため、文書の意味に係わるような分析作業において、単にその結果のみを出力するのではなく、情報分析作業全般にわたる支援をおこなうことが可能な文書処理装置が得られるという効果を奏する。

#### 【0116】

また、請求項2の発明によれば、前記出力手段が、前記加工処理手段により加工処理された文書データの全部または一部の内容に基づいて複数の項目値を設定する項目値設定手段と、前記項目値設定手段により設定された項目値ごとに前記文書データの全部または一部を集計する集計手段と、を備え、前記文書データの全部または一部を、項目値を少なくとも1つの軸とする表形式に展開して出力するため、簡易な操作で加工処理の結果をクロス表として表すことができ、情報の内容の把握を容易におこなうことができることから、文書の意味に係わるような分析作業において、単にその結果のみを出力するのではなく、情報分析作業全般にわたる支援をおこなうことが可能な文書処理装置が得られるという効果を奏する。

#### 【0117】

また、請求項3の発明によれば、前記出力手段が、さらに、前記加工処理手段により加工処理された文書データの全部または一部を、前記加工処理手段により加工処理される前の文書データの全部または一部とともに出力するため、加工処理すべき対象データとその他のデータが同時に表示され、それを確認することにより、加工処理の対象範囲の決定を正確かつ容易におこなうことができることから、文書の意味に係わるような分析作業において、単にその結果のみを出力するのではなく、情報分析作業全般にわたる支援をおこなうことが可能な文書処理装置が得られるという効果を奏する。

#### 【0118】

また、請求項4の発明によれば、前記文書記憶手段が、さらに、前記加工処理手段により加工処理された文書データの全部または一部を記憶するため、以後、他のデータと同様に扱うことができることから、文書の意味に係わるような分析

作業において、単にその結果のみを出力するのではなく、情報分析作業全般にわたる支援をおこなうことが可能な文書処理装置が得られるという効果を奏する。

#### 【0119】

また、請求項5の発明によれば、前記選択手段が、さらに、前記出力手段により出力された文書データの全部または一部を選択するため、出力手段により出力された文書データの全部又は一部をさらなる分析の対象とすることができます、多彩で高度な情報分析作業ができることから、文書の意味に係わるような分析作業において、単にその結果のみを出力するのではなく、情報分析作業全般にわたる支援をおこなうことが可能な文書処理装置が得られるという効果を奏する。

#### 【0120】

また、請求項6の発明によれば、前記文書記憶手段が、さらに、前記加工処理の内容に関するデータを記憶するため、加工処理の内容に関するデータの紛失を防止し、当該データの管理が容易になるだけでなく、加工処理にもちいた設定とそれによる処理結果を関連づけて把握することができることから、文書の意味に係わるような分析作業において、単にその結果のみを出力するのではなく、情報分析作業全般にわたる支援をおこなうことが可能な文書処理装置が得られるという効果を奏する。

#### 【0121】

また、請求項7の発明によれば、入力された文書データを記憶する文書記憶工程と、前記文書記憶工程により記憶された文書データの全部または一部を選択する選択工程と、前記選択工程により選択された文書データの全部または一部から文字列の特徴に関するデータを抽出する特徴抽出工程と、前記特徴抽出工程により抽出された文字列の特徴に関するデータに基づいて前記文書データの全部または一部を加工処理する加工処理工程と、前記加工処理工程により加工処理された文書データの全部または一部を出力する出力工程と、を含むので、文書の意味に係わるような分析作業において、単にその結果のみを出力するのではなく、情報分析作業全般にわたる支援をおこなうことが可能な文書処理方法が得られるという効果を奏する。

#### 【0122】

また、請求項8の発明によれば、前記出力工程が、前記加工処理工程により加工処理された文書データの全部または一部の内容に基づいて複数の項目値を設定する項目値設定工程と、前記項目値設定工程により設定された項目値ごとに前記文書データの全部または一部を集計する集計工程と、を含み、前記文書データの全部または一部を、項目値を少なくとも1つの軸とする表形式に展開して出力するので、簡易な操作で加工処理の結果をクロス表として表すことができ、情報の内容の把握を容易におこなうことができることから、文書の意味に係わるような分析作業において、単にその結果のみを出力するのではなく、情報分析作業全般にわたる支援をおこなうことが可能な文書処理方法が得られるという効果を奏する。

#### 【0123】

また、請求項9の発明によれば、前記出力工程が、さらに、前記加工処理工程により加工処理された文書データの全部または一部を、前記加工処理工程により加工処理される前の文書データの全部または一部とともに出力するので、加工処理すべき対象データとその他のデータが同時に表示され、それを確認することにより、加工処理の対象範囲の決定を正確かつ容易におこなうことができることから、文書の意味に係わるような分析作業において、単にその結果のみを出力するのではなく、情報分析作業全般にわたる支援をおこなうことが可能な文書処理方法が得られるという効果を奏する。

#### 【0124】

また、請求項10発明によれば、前記文書記憶工程が、さらに、前記加工処理工程により加工処理された文書データの全部または一部を記憶するので、以後、他のデータと同様に扱うことができることから、文書の意味に係わるような分析作業において、単にその結果のみを出力するのではなく、情報分析作業全般にわたる支援をおこなうことが可能な文書処理方法が得られるという効果を奏する。

#### 【0125】

また、請求項11の発明によれば、前記選択工程が、さらに、前記出力工程により出力された文書データの全部または一部を選択するので、出力工程により出力された文書データの全部又は一部をさらなる分析の対象とすることができます、多

高度な情報分析作業ができることから、文書の意味に係わるような分析作業において、単にその結果のみを出力するのではなく、情報分析作業全般にわたる支援をおこなうことが可能な文書処理方法が得られるという効果を奏する。

## 【0126】

また、請求項12の発明によれば、前記文書記憶工程が、さらに、前記加工處理の内容に関するデータを記憶するので、加工処理の内容に関するデータの紛失を防止し、当該データの管理が容易になるだけでなく、加工処理にもちいた設定とそれによる処理結果を関連づけて把握することができることから、文書の意味に係わるような分析作業において、単にその結果のみを出力するのではなく、情報分析作業全般にわたる支援をおこなうことが可能な文書処理方法が得られるという効果を奏する。

## 【0127】

また、請求項13の発明によれば、請求項7~12のいずれか一つに記載された方法をコンピュータに実行させるプログラムを記録したことで、そのプログラムを機械読み取り可能となり、これによって、請求項7~12の動作をコンピュータによって実現することが可能な記録媒体が得られるという効果を奏する。

## 【図面の簡単な説明】

## 【図1】

この発明の本実施の形態による文書処理装置を構成する情報処理システム全体のハードウェア構成を示す説明図である。

## 【図2】

本実施の形態による文書処理装置を構成する情報処理システムにおけるサーバーをハードウェア的に示す説明図である。

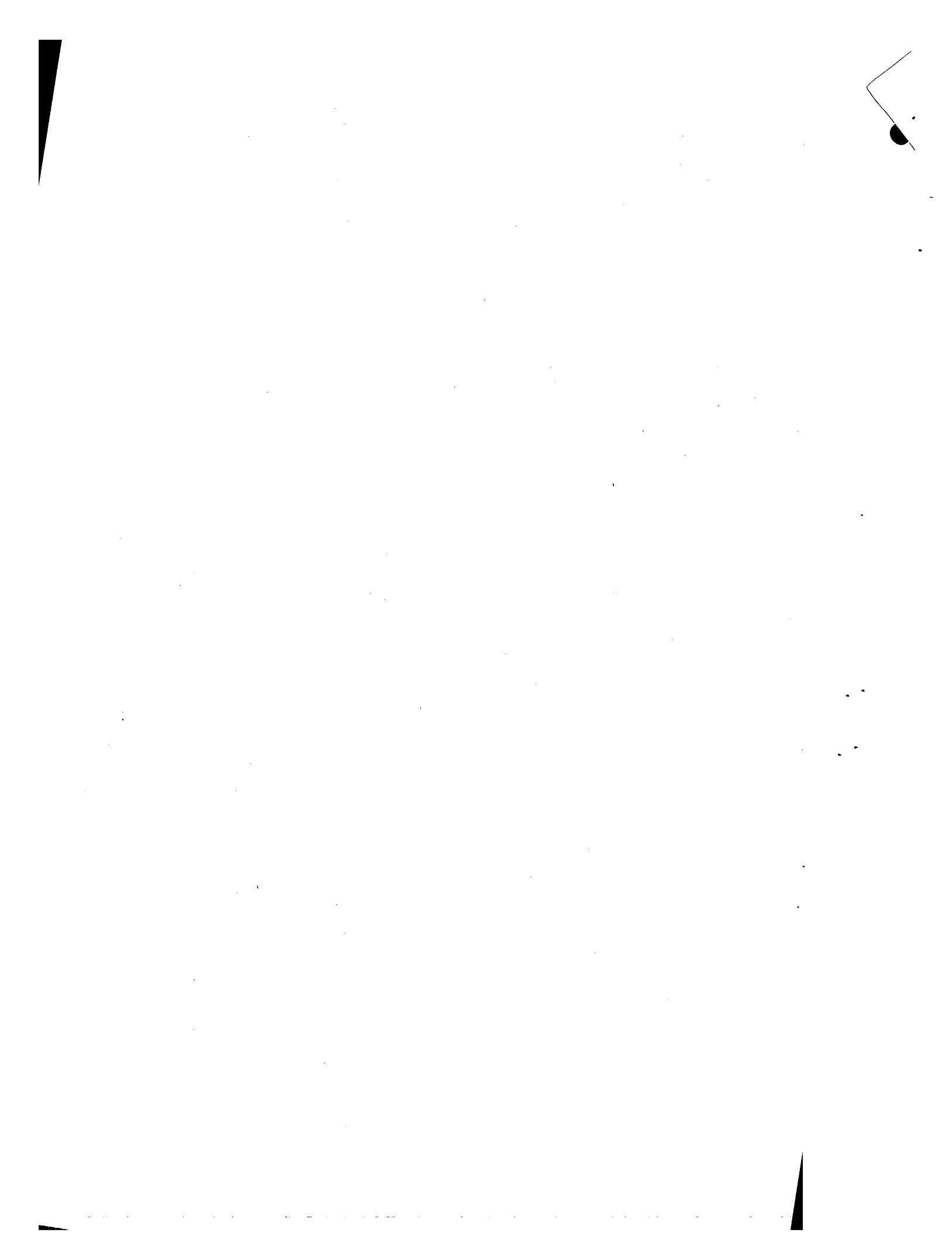
## 【図3】

本実施の形態による文書処理装置を構成する情報処理システムにおけるクライアントをハードウェア的に示す説明図である。

## 【図4】

本実施の形態による文書処理装置の構成を機能的に示すブロック図である。

## 【図5】





本実施の形態による文書処理装置の項目名と項目値の関係を示す説明図である

【図6】

本実施の形態による文書処理装置の文書記憶部に記憶された文書のデータ構造を示す説明図である。

【図7】

本実施の形態による文書処理装置の文書記憶部に記憶された文書の別のデータ構造を示す説明図である。

【図8】

実施の形態による文書処理装置の出力部による画面表示の例を示す説明図である。

【図9】

実施の形態による文書処理装置の出力部による画面表示の別の例を示す説明図である。

【図10】

実施の形態による文書処理装置の出力部による画面表示の別の例を示す説明図である。

【図11】

本実施の形態による文書処理装置の特徴抽出部によりおこなわれる抽出処理の内容の一覧を示す説明図である。

【図12】

本実施の形態による文書処理装置の加工処理部によりおこなわれる加工処理の内容の一覧を示す説明図である。

【図13】

本実施の形態による文書処理装置の各項目の特徴ベクトルを示す説明図である

【図14】

本実施の形態による文書処理装置の単語とその単語IDごとの出現回数を示す説明図である。

**【図15】**

本実施の形態による文書処理装置の出力部による画面表示の別の例を示す説明図である。

**【図16】**

本実施の形態による文書処理装置の出力部によるクロス表作成ための指示画面を示す説明図である。

**【図17】**

本実施の形態による文書処理装置の出力部による分類処理の結果が表示されたクロス表を示す説明図である。

**【図18】**

本実施の形態による文書処理装置の出力部による分類処理の結果が表示された別のクロス表を示す説明図である。

**【図19】**

本実施の形態による文書処理装置の出力部の詳細な構成を示すブロック図である。

**【図20】**

本実施の形態による文書処理装置のクロス表の出力手順を示すフローチャートである。

**【図21】**

本実施の形態による文書処理装置の出力部による画面表示の別の例を示す説明図である。

**【図22】**

本実施の形態による文書処理装置の出力部による画面表示の別の例を示す説明図である。

**【図23】**

本実施の形態による文書処理装置の出力部による画面表示の別の例を示す説明図である。

**【図24】**

本実施の形態による文書処理装置の出力部による画面表示の別の例を示す説明

図である。

## 【図25】

本実施の形態による文書処理装置の文書記憶部の詳細な構成を示すブロック図である。

## 【図26】

本実施の形態による文書処理装置の出力部による画面表示の別の例を示す説明図である。

## 【図27】

本実施の形態による文書処理装置の出力部による画面表示の別の例を示す説明図である。

## 【図28】

本実施の形態による文書処理装置の出力部による画面表示の別の例を示す説明図である。

## 【図29】

本実施の形態による文書処理装置の文書処理の一連の手順を示すフローチャートである。

## 【符号の説明】

101 サーバー

102 クライアント

103 ネットワーク

201 CPU

204 I/F

206 ディスク装置

301 CPU

306 ハードディスク

308 ディスプレイ

309 I/F

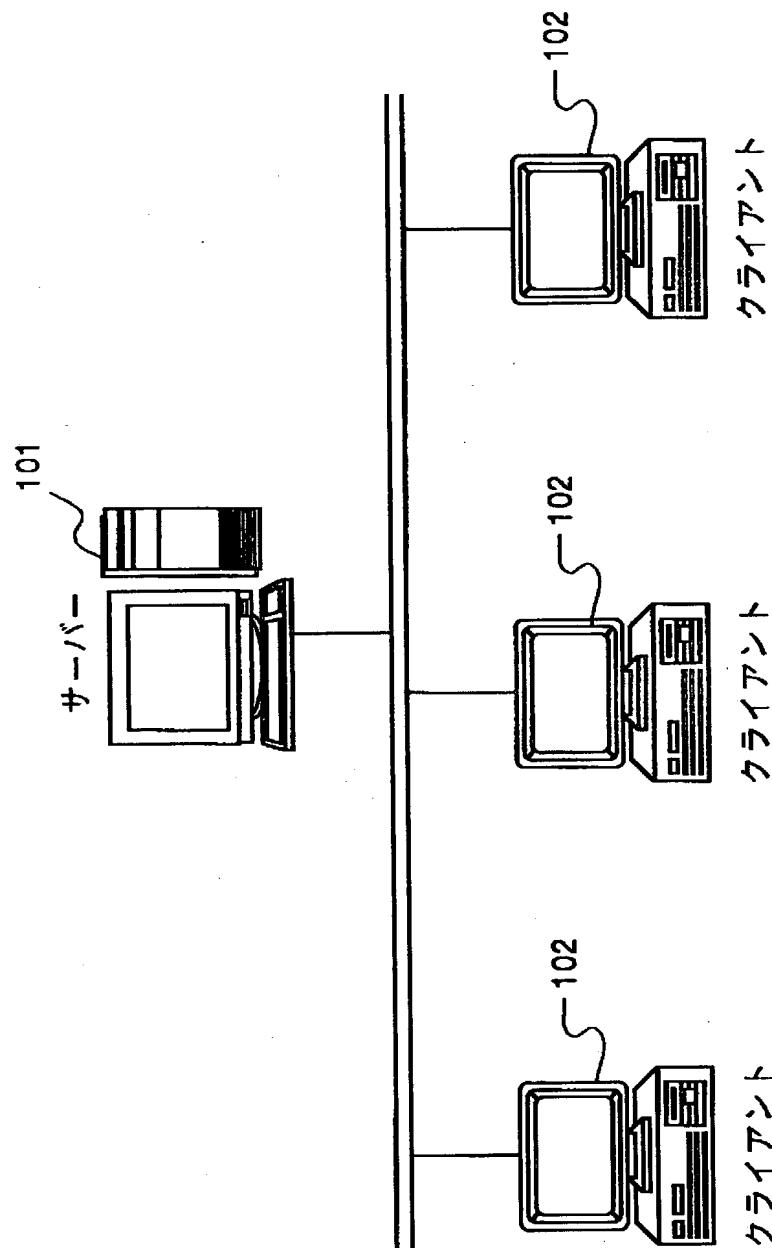
311 キーボード

312 マウス

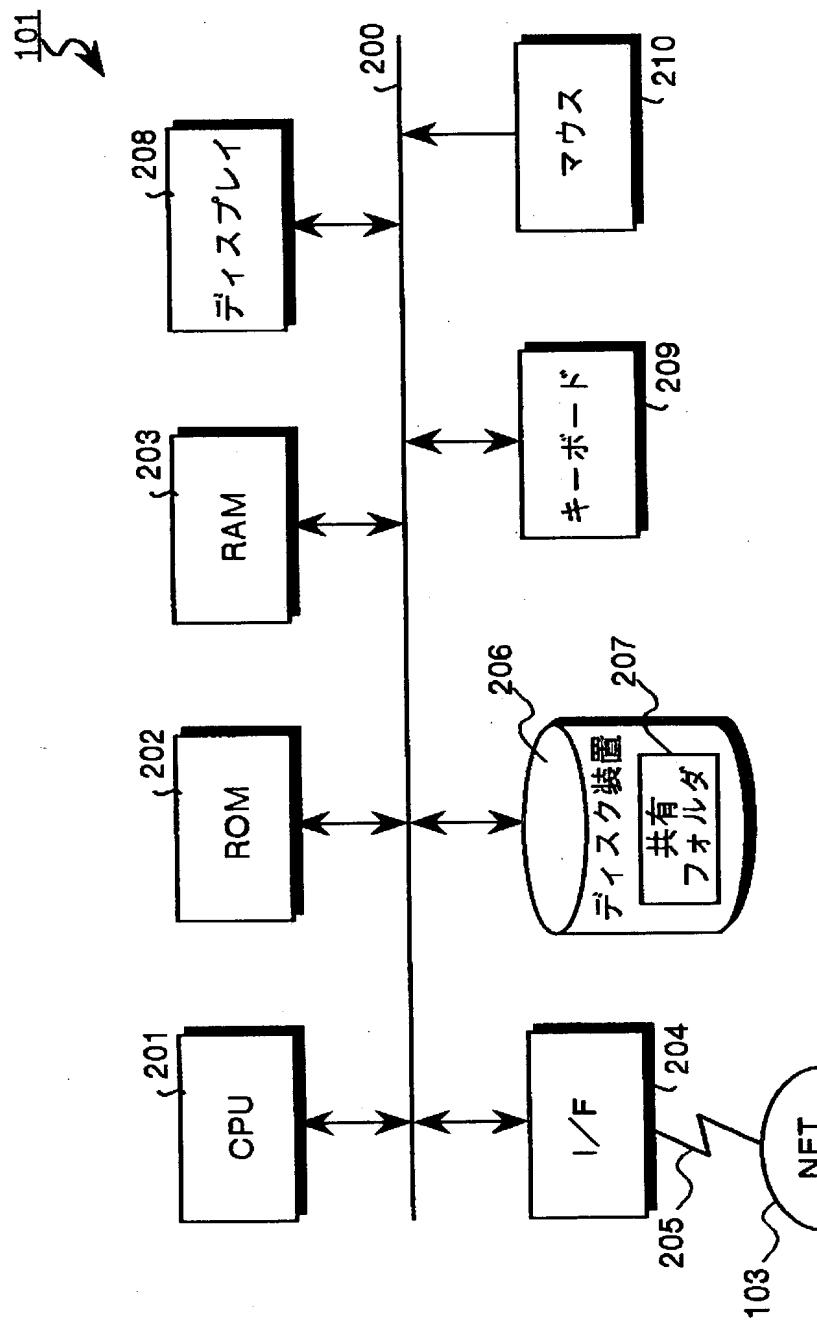
- 313 スキャナ
- 400 バス
- 401 入力部
- 402 文書記憶部
- 403 選択部
- 404 特徴抽出部
- 405 加工処理部
- 406 出力部
- 407 グラフ描画部
- 408 加工処理結果保持部
- 409 解析部
- 410 特徴ベクトル生成部
- 1600 指示画面
- 1700 クロス表
- 1800 マウスポインタ
- 1801 内容表示画面
- 1901 項目値選定部
- 1902 集計部
- 1903 表保持部
- 2401 棒グラフ表示領域
- 2501 設定値記憶部
- 2502 設定値送受信部
- 2503 分類情報記憶部
- 2604 問い合わせ画面
- 2703 分類情報表示画面
- 2802 表示領域

【書類名】 図面

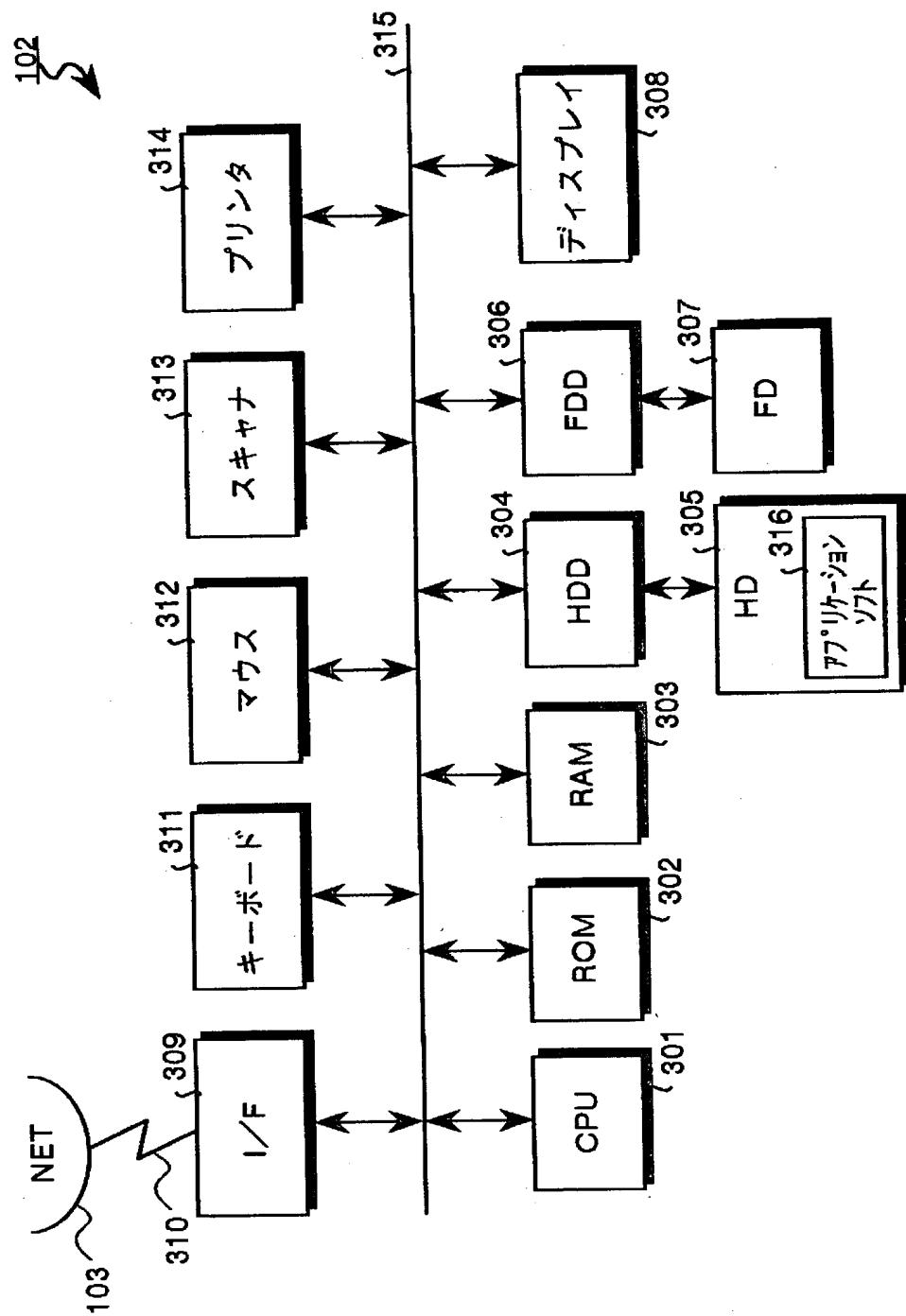
【図1】



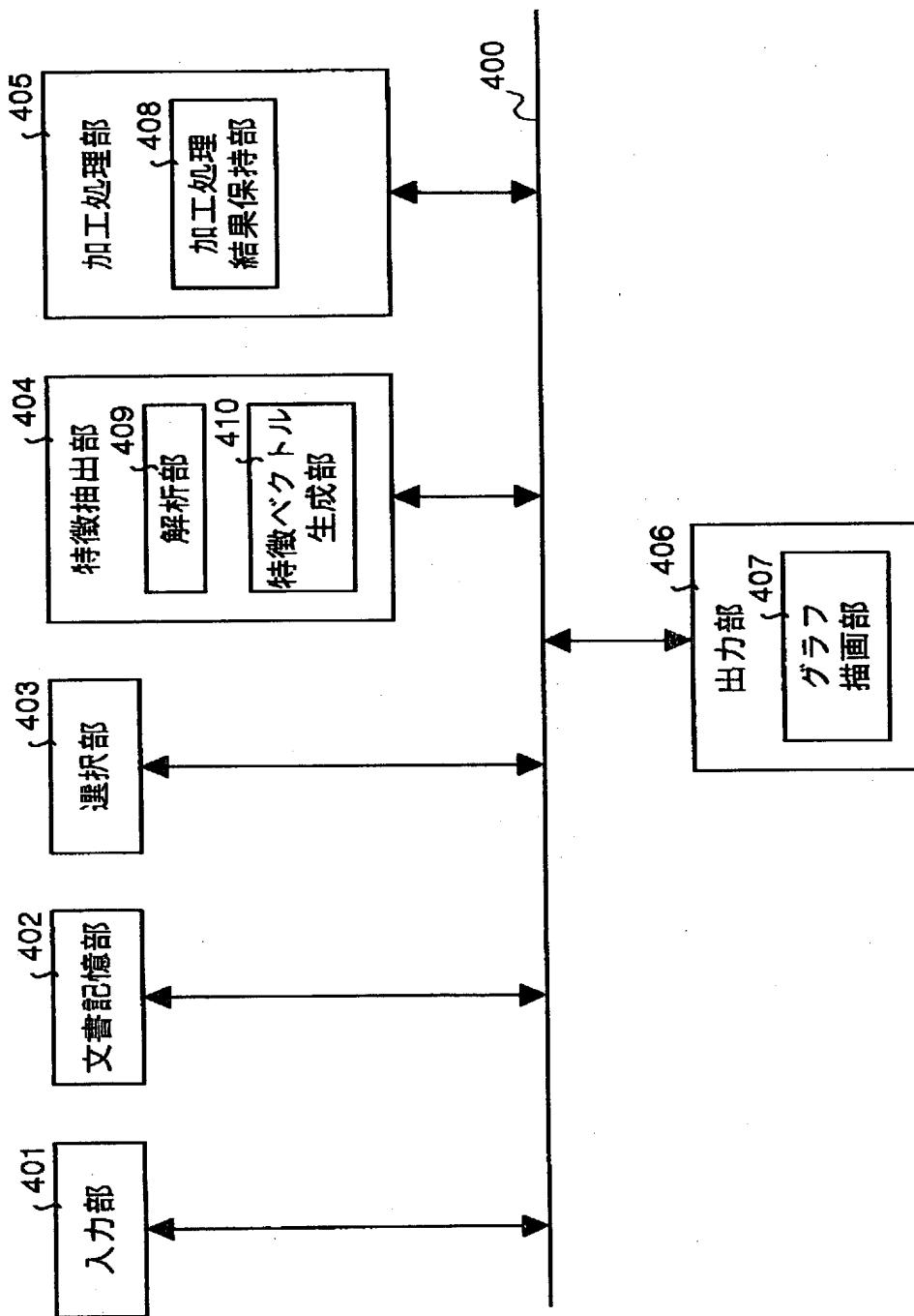
【図2】



【図3】



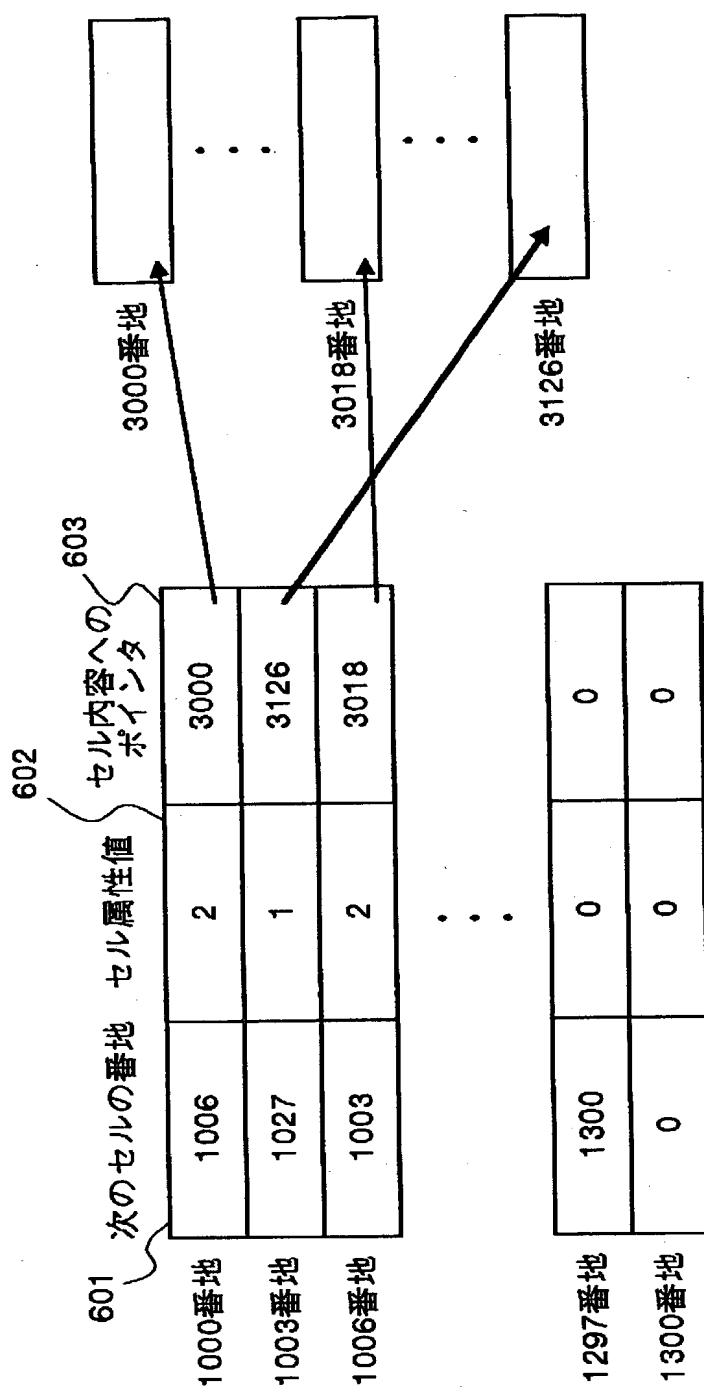
【図4】



【図5】

項目名	項目値
出願番号	特願平10-00000
出願日	平成10年1月1日
請求項の数	3
発明の名称	画像形成装置
発明の効果	本発明の画像形成装置は、○○を××しているので…

【図6】



【図7】

701 セル属性値 セル内容への  
702 ポイント

1000番地	2	3000	2	3018
1100番地	1	3421	1	3512
				..
1500番地	0	0	0	0
				..

【図8】

番号	受付日	業所	車種	年式	内容	
					801	802
1	1997/3/5	愛知	ABC1800	1993	騒音が大きい	△
2	1997/3/5	電気	ABC2000	1995	排氣が黒い	
3	1997/3/5	東京	ABC1800	1996	塗装が変色する	
4	1997/3/5	幌岡	DEF1600	1995	オイルが漏れる	
5	1997/3/5	別	KLM1200	1992	暖房が効かない	
6	1997/3/5	福登	DEF1600	1994	騒音が大きい	
7	1997/3/5	長野	DEF1600	1996	エンジンがかからない	
8	1997/3/5	東京	ABC1800	1997	オーバーヒートが起かる	
9	1997/3/6	高松	XYZ3000	1992	バッテリーが上がる	
10	1997/3/6	崎	KLM1200	1993	エンジンがかからない	
11	1997/3/6	大阪	ABC1600	1994	排氣が黒い	
12	1997/3/6	長野	DEF1600	1997	ラジオが鳴らない	
13	1997/3/6	岡	ABC1800	1996	塗装がはげる	
14	1997/3/6	仙台	XYZ3000	1995	暖房が効かない	

【図9】

番号	受付日	年式	車種	内容
11	1997/3/6	1994	ABC1600	排気が黒い
53	1997/4/21	1993	ABC1600	オイルが漏れる
1	1997/3/5	1993	ABC1800	排気音が大きくなる
3	1997/3/5	1996	ABC1800	オイル漏れ
8	1997/3/5	1997	ABC1800	オイル漏れ
13	1997/3/6	1996	ABC1800	オイル漏れ
18	1997/3/10	1994	ABC1800	オイル漏れ
28	1997/3/12	1995	ABC1800	オイル漏れ
35	1997/3/17	1996	ABC1800	オイル漏れ
39	1997/3/20	1993	ABC1800	オイル漏れ
42	1997/3/22	1996	ABC1800	オイル漏れ
46	1997/3/24	1997	ABC1800	オイル漏れ
47	1997/3/24	1994	ABC1800	オイル漏れ
4	1997/3/5	1995	DEF1600	オイル漏れ

【図10】

1001

		内客		年式		車種		営業所		受付日		番号	
	A	徐氣イ音表一達才塗	ガ事が漏れる	1994	ABC1600	大阪	長野	京都	岡山	福島	1997/3/6	11	
		才塗	大塗	1993	ABC1600	東京	東京	東京	大阪	鳥取	1997/4/21	53	
		塗	大塗	1993	ABC1800	東京	東京	東京	大阪	滋賀	1997/3/5	1	
		塗	大塗	1996	ABC1800	大阪	大阪	大阪	大阪	滋賀	1997/3/5	3	
		塗	大塗	1997	ABC1800	大阪	大阪	大阪	大阪	滋賀	1997/3/5	8	
		塗	大塗	1996	ABC1800	大阪	大阪	大阪	大阪	滋賀	1997/3/6	13	
		塗	大塗	1994	ABC1800	大阪	大阪	大阪	大阪	滋賀	1997/3/10	18	
		塗	大塗	1995	ABC1800	大阪	大阪	大阪	大阪	滋賀	1997/3/12	28	
		塗	大塗	1996	ABC1800	大阪	大阪	大阪	大阪	滋賀	1997/3/17	35	
		塗	大塗	1993	ABC1800	大阪	大阪	大阪	大阪	滋賀	1997/3/20	39	
		塗	大塗	1996	ABC1800	大阪	大阪	大阪	大阪	滋賀	1997/3/22	42	
		塗	大塗	1997	ABC1800	大阪	大阪	大阪	大阪	滋賀	1997/3/24	46	
		塗	大塗	1994	ABC1800	大阪	大阪	大阪	大阪	滋賀	1997/3/24	47	
		塗	大塗	1995	DEF1600	大阪	大阪	大阪	大阪	滋賀	1997/3/5	4	

【図11】

抽出処理内容	
1	対象とする文字列に含まれる単語
2	対象とする文字列に含まれる単語数
3	対象とする文字列に含まれる単語の文字数
4	対象とする文字列に含まれる単語それぞれの出現回数
5	対象とする文字列に含まれる単語それぞれの品詞
6	対象とする文字列に含まれる単語間の関係の情報
7	対象とする文字列に含まれる文の数
8	対象とする文字列に含まれる文の文字数
9	対象とする文字列に含まれる文の文節数
10	対象とする文字列に含まれる文の間の関係
:	:

【図12】

加工処理内容	
1	分類処理
2	検索処理
3	並べ替え処理
4	代表値抽出処理
5	最大値抽出処理
6	最小値抽出処理
7	算術処理
:	:

【図15】

▲ ↗

1501 ↘

番号	受付日	営業所	車種	年式	内容	カラーバリ番号
1	1997/3/5	愛知	ABC1800	1993	騒音が大きい	5
2	1997/3/5	東京	ABC2000	1995	騒音が黒い	1
3	1997/3/5	札幌	ABC1800	1996	排気音が変色する	7
4	1997/3/5	福岡	DEF1600	1995	塗装が漏れる	11
5	1997/3/5	別府	KLM1200	1992	オイルが効かない	2
6	1997/3/5	長野	DEF1600	1994	騒音が大きい	5
7	1997/3/5	東京	DEF1600	1996	エンジンがかからない	8
8	1997/3/5	松山	ABC1800	1997	オーバーヒートが起る	14
9	1997/3/6	高崎	XYZ3000	1992	バッテリーが上がる	12
10	1997/3/6	阪神	KLM1200	1993	エンジンが黒い	8
11	1997/3/6	大阪	ABC1600	1994	排気が鳴らない	1
12	1997/3/6	長野	DEF1600	1997	ラジオが鳴らない	6
13	1997/3/6	盛岡	ABC1800	1996	塗装がはげる	7
14	1997/3/6	仙台	XYZ3000	1995	暖房が効かない	2 ▽

↖

【図16】

1600

↓

<クロス表の作成>	
処理対象項目名：	(1601) 内容
軸となる項目名：	(1602) 車種
分類数：	(1603) 50
縦軸	横軸
(1604)	(1605)

【図17】

1700 ↗

車種	ABC1600	ABC1800	ABC2000	DEF1600	KLM1200	XYZ3000	合計
クラス1	3	0	4	1	1	0	9
クラス2	0	0	2	0	0	23	25
クラス3	5	3	2	7	0	4	21
クラス4	7	6	7	0	0	1	21
.....	.....	.....	.....	.....	.....	.....	.....
合計	227	135	87	194	134	281	1058

【図18】

車種	ABC1600	ABC1800	ABC2000	DEF1600	KLM1200	XYZ3000	合計
クラス1	3	0	4	1	1	0	9
クラス2				0	0	23	25
クラス3				7	0	4	21
クラス4					0	1	1
.....				.....	.....	.....	.....
合	194	134	281	1058			

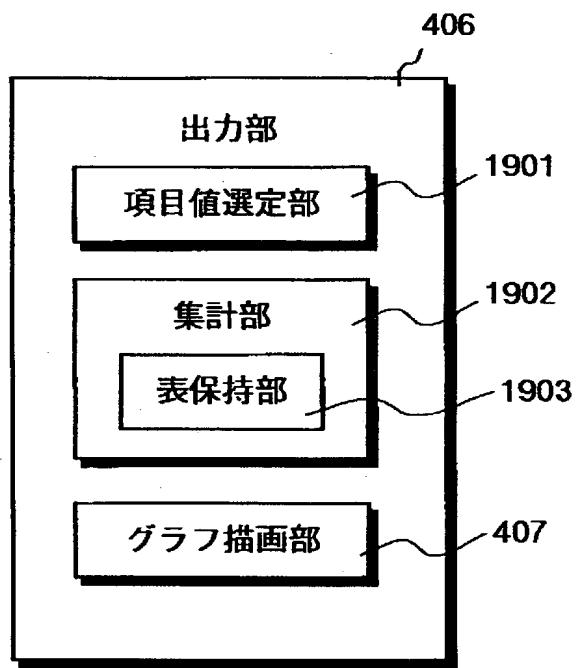
1800

1700

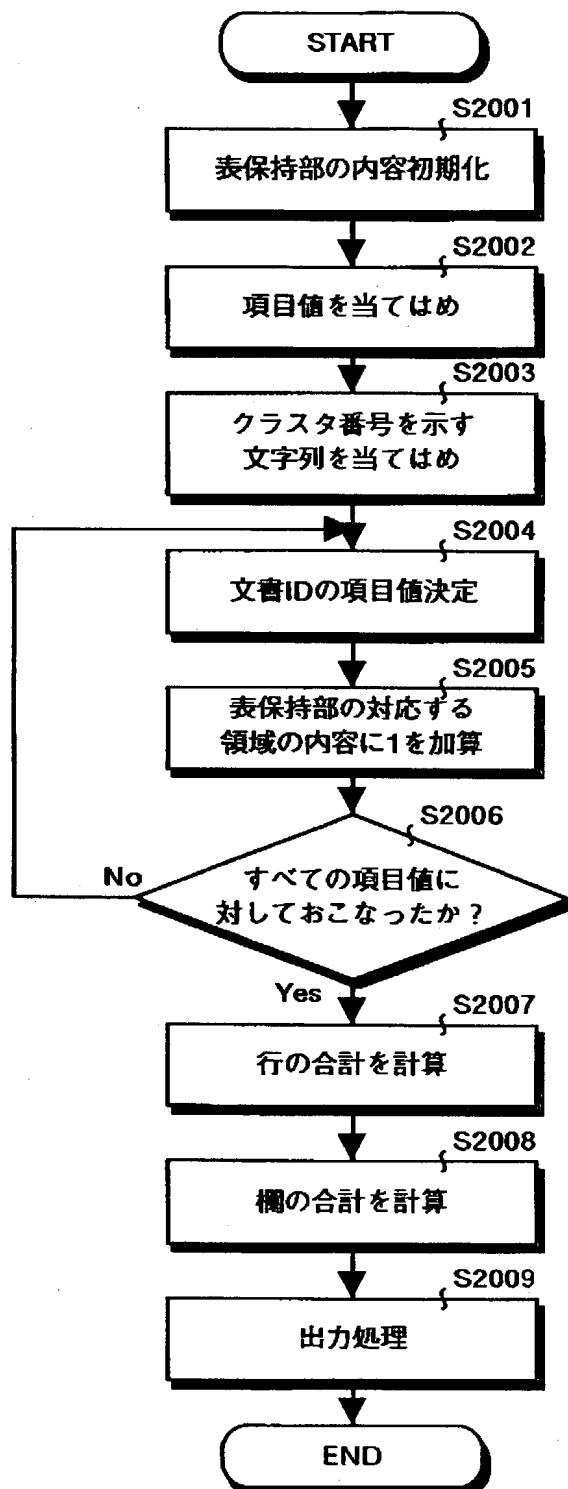
1801

データ数： 4  
表示項目： 内容  
セル： ABC2000-クラスタ1  
.....  
排気が黒い  
排気が黒い  
排気が黒い  
排気が黒い

【図19】



【図20】



【図21】

2101

番号	受付日	業者所	年式	車種	内容
1	1997/3/5	愛知	1993	ABC1800	騒音が大きい
2	1997/3/5	富山	1995	ABC2000	排気ガス漏れる
3	1997/3/5	京	1996	ABC1800	塗装が変色する
4	1997/3/5	幌	1996	DEF1600	才イルが効かない
5	1997/3/5	岡	1995	KLM1200	暖房音が大きい
6	1997/3/5	別	1992	DEF1600	エンジンがかからない
7	1997/3/5	長	1994	DEF1600	オーバーヒートが起る
8	1997/3/5	野	1994	ABC1800	バッテリーが上がる
9	1997/3/6	東	1996	XYZ3000	エンジンが黒い
10	1997/3/6	福	1992	KLM1200	排気が漏らない
11	1997/3/6	登	1993	ABC1600	ラジオが鳴らない
12	1997/3/6	長	1994	DEF1600	塗装がはげる
13	1997/3/6	高	1996	ABC1800	暖房が効かない
14	1997/3/6	嶺	1995	XYZ3000	仙台

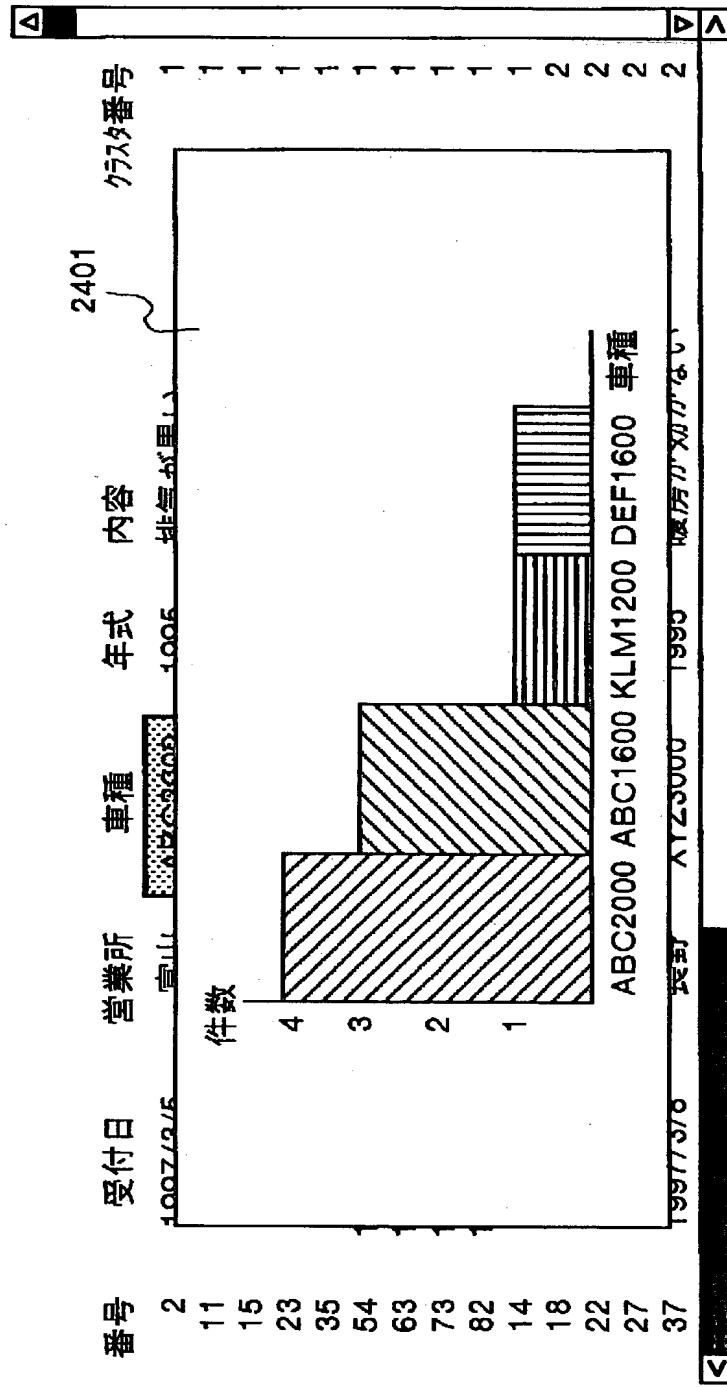
【図22】

カラタ番号	内容	年式	車種	営業所	受付日	番号
1	排気が黒い	1995	ABC2000	電山	1997/3/5	2
1	排気が黒い	1994	ABC1600	東京	1997/3/6	11
1	排気が黒い	1996	ABC2000	幌岡	1997/3/7	15
1	排気が黒い	1995	ABC2000	長野	1997/3/7	23
1	排気がくさい	1992	KLM1200	東京	1997/3/8	35
1	排気が黒い	1994	ABC1600	福岡	1997/3/10	54
1	排気が黒い	1996	ABC2000	長野	1997/3/12	63
1	排気が黒い	1997	DEF1600	東京	1997/3/14	73
1	排気が黒い	1992	ABC1600	岡谷	1997/3/14	82
1	排気が効かない	1995	XYZ3000	長野	1997/3/6	14
1	暖房が効かない	1995	XYZ3000	仙台	1997/3/7	18
1	暖房が効かない	1997	XYZ3000	長野	1997/3/7	22
2	暖房の効きが悪い	1995	XYZ3000	台	1997/3/8	27
2	暖房が効かない	1995	XYZ3000	長野	1997/3/8	37

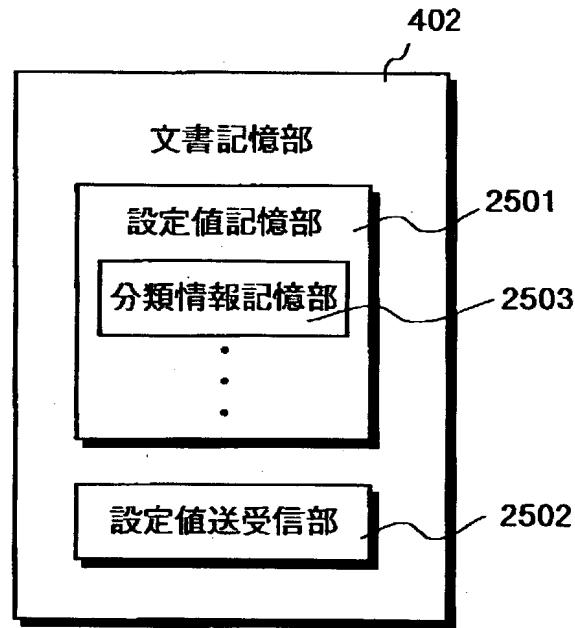
【図23】

2301		番号	受付日	営業所	車種	年式	内容	クラス番号
2	1997/3/5	ABC2000	1995	山東長野	排気が黒い	1	1	1
11	1997/3/6	ABC1800	1994	京福長野	排気が黒い	1	1	1
15	1997/3/7	ABC2000	1996	京福長野	排気が黒い	1	1	1
23	1997/3/7	ABC2000	1995	京福長野	排気が黒い	1	1	1
35	1997/3/8	KLM1200	1992	京福長野	排気が黒い	1	1	1
54	1997/3/10	ABC1600	1994	仙台	排気が黒い	1	1	1
63	1997/3/12	ABC2000	1996	仙台	排気が黒い	1	1	1
73	1997/3/14	DEF1800	1997	仙台	排気が黒い	1	1	1
82	1997/3/14	ABC1600	1992	仙台	排気が黒い	1	1	1
14	1997/3/6	XYZ3000	1995	長野	暖房が効かない	1	1	1
18	1997/3/7	XYZ3000	1997	長野	暖房が効かず、寒い	2	2	2
22	1997/3/7	XYZ3000	1997	長野	暖房が効かない	2	2	2
27	1997/3/8	XYZ3000	1995	長野	暖房の効きが悪い	2	2	2
37	1997/3/8	XYZ3000	1995	長野	暖房が効かない	2	2	2

### 【図24】



【図25】



【図26】

番号	登仕日	登業証	分類	直種	年式
2605	1997/3/6	50	11	分類数を入力してください	1994
2603	1997/3/10		53	盛岡	1993
2602	1997/3/12		1	大阪	1993
2604	1997/3/17		3	広島	1996
2601	1997/3/20		8	横浜	1997
	1997/3/22		13	知	1996
	1997/3/24		18	東京	1996
	1997/3/24		28	富山	1994
	1997/3/24		35	大阪	1995
	1997/3/5		39	札幌	1996
			42		1993
			46		1996
			47		1997
			4		1994
					1995

【図27】

車種			年式	内容	△
ABC18	ABC20	ABC18	1997/3/5	分類日時： 1997年5月12日 16時32分	2701
DEF18	KLM12	DEF16	1997/3/5	分類対象数： 823	2703
DEF16	DEF16	ABC18	1997/3/5	<分類設定値>	2702
ABC18	XYZ30	XYZ30	1997/3/5	分類数： 30	▲
XYZ30	KLM1200	XYZ30	1997/3/6	分類品詞： 名詞	12
KLM1200	ABC1600	ABC1600	1997/3/6	1993 上ノゾムががちない	8
ABC1600	DEF1600	ABC1800	1997/3/6	排氣が黒い ラジオが鳴らない	1 6
DEF1600	ABC1800	XYZ3000	1997/3/6	塗装がはげる 暖房が効かない	7 2 ▽
ABC1800	XYZ3000		1997/3/6		▲
XYZ3000			1997/3/6		▼

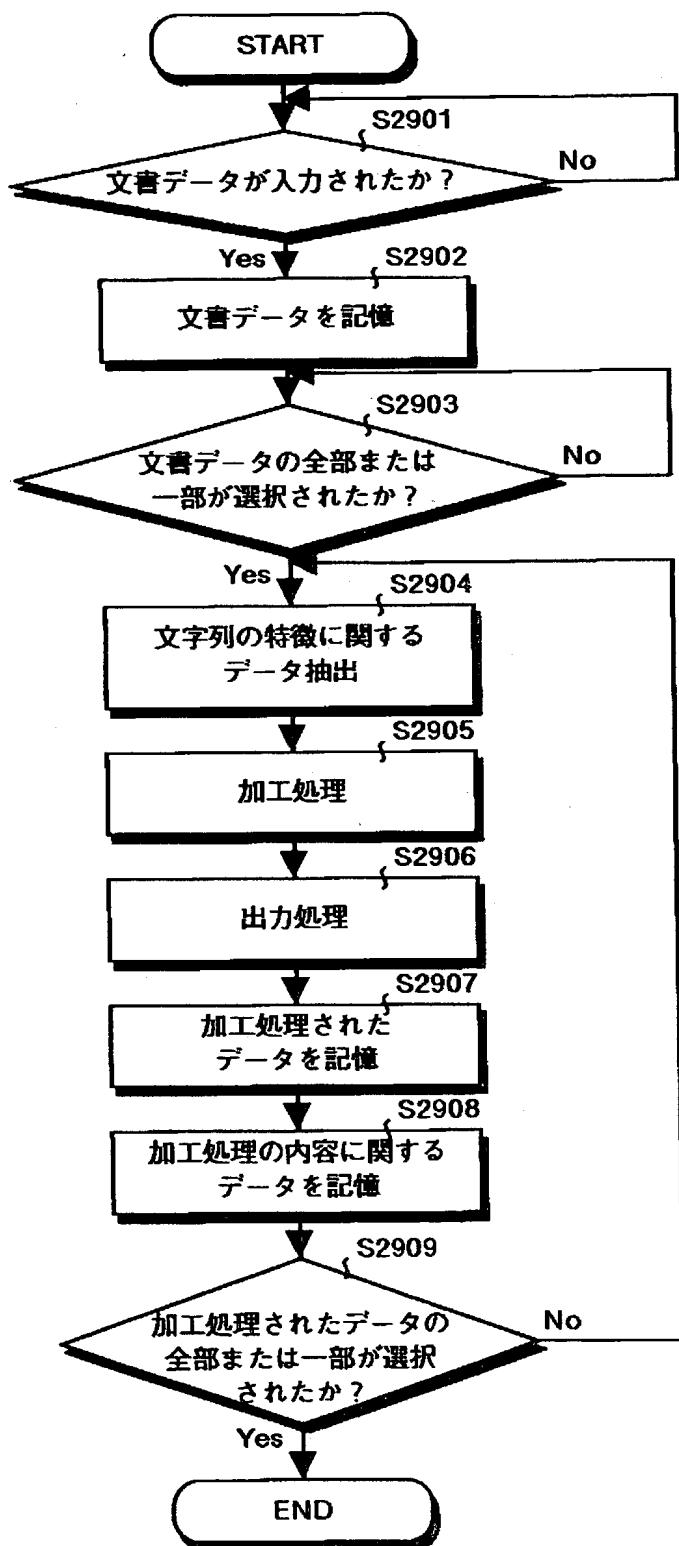
【図28】

番号	受付日	営業所	車種	年式	内容	クラス番号	
						△	▽
1	1997/3/5	愛知	ABC1800	1993	騒音が大きい	5	1
2	1997/3/5	東京	ABC2000	1995	排気が黒い	1	7
3	1997/3/5	札幌	ABC1800	1996	塗装が変色する	7	11
4	1997/3/5	別府	DEF1600	1995	オイルが漏れる	11	2
5	1997/3/5	長野	KLM1200	1992	暖房が効かない	2	5
6	1997/3/5	東京	DEF1600	1994	騒音が大きい	5	8
7	1997/3/5	福島	DEF1600	1996	エンジンがかからない	8	14
8	1997/3/5	長野	ABC1800	1997	オーバーヒートが起まる	14	12
9	1997/3/6	東京	XYZ3000	1992	バッテリーが上がる	12	8
10	1997/3/6	高松	KLM1200	1993	エンジンがかからない	8	1
11	1997/3/6	大阪	ABC1600	1994	排気が黒い	1	6
12	1997/3/6	長野	DEF1600	1997	ラジオが鳴らない	6	7
13	1997/3/6	岡山	ABC1800	1996	塗装がはげる	7	2
14	1997/3/6	仙台	XYZ3000	1995	暖房が効かない	2	2801

&lt;分類設定値&gt;分類数：30 分類品詞：名詞

分類結果1 分類結果2

【図29】



【書類名】 要約書

【要約】

【課題】 文書の意味に係わるような分析作業において、単にその結果のみを出力するのではなく、情報分析作業全般にわたる支援をおこなうことを課題とする。

【解決手段】 入力された文書データを記憶する文書記憶部402と、文書記憶部402により記憶された文書データの全部または一部を選択する選択部403と、選択部403により選択された文書データの全部または一部から文字列の特徴に関するデータを抽出する特徴抽出部404と、特徴抽出部404により抽出された文字列の特徴に関するデータに基づいて文書データの全部または一部を加工処理する加工処理部405と、加工処理部405により加工処理された文書データの全部または一部を出力する出力部406とを備える。

【選択図】 図4

出願人履歴情報

識別番号 [000006747]

1. 変更年月日 1990年 8月24日

[変更理由] 新規登録

住 所 東京都大田区中馬込1丁目3番6号

氏 名 株式会社リコー